Title
A description-experience gap in face stereotyping

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Authors
Wang, Hongyi
Lu, Yilun
He, Lisheng

Publication Date
2023

Peer reviewed

A description-experience gap in face stereotyping

Hongyi Wang (hywang@ecnu.edu.cn)
School of Psychology and Cognitive Science, East China Normal University
Shanghai, China

Yilun Lu (yilun0519@126.com)
Shanghai Jiao Tong University & East China Normal University
Shanghai, China

Lisheng He (hlisheng@shu.edu.cn) [1]
SILC Business School, Shanghai University
Shanghai, China

Abstract

Face stereotypes are prevalent and consequential. In this paper, we investigate an experience sampling account of face stereotyping. In two experiments, we show that, in an experience-sampling-based Trust Game, participants in the role of an investor were much more likely to play the game with trustees with trustworthy faces than those with untrustworthy faces (Experiment 1). Crucially, this endogenous experience sampling bias has amplified the behavioral consequences of the facial trustworthiness stereotype. In contrast, when the information was directly described with no sampling constraint for the participants, the face stereotype had a very limited effect on investment decisions (Experiment 2). We conclude by suggesting that the description-experience gap paves a promising avenue for studying sampling-based accounts in social cognition and behavior.

Keywords: Face stereotype; description-experience gap; information sampling; trust game; eye tracking

Introduction

Human faces are an important source of stereotypes. They contain rich high-dimensional information about individuals. When encountering a stranger, people extract not only basic categorical information, such as sex, age, and ethnicity, but also rapidly infer trait judgments from facial appearances (Bar et al., 2006; Willis & Todorov, 2006). For example, people tend to attribute positive qualities (e.g., trustworthiness, intelligence, health) to attractive people, which is known as the "what is beautiful is good" stereotype (Dion, Berscheid, & Walster, 1972). These facial stereotypes can be formed quickly and prevalent in everyday life, such as in schools and labor markets (see Maestripieri et al., 2017 for a review).

Stereotypical evaluations can have substantial effects on people's behavior. For instance, people prefer to date with, hire, and vote for attractive individuals (Langlois et al., 2000). In addition, the perceived trustworthiness of a person's face influences the extent to which others cooperate with them in socioeconomic interactions (Van't Wout & Sanfey, 2008). The effect of perceived facial trustworthiness on cooperative behavior is even evident in children as young as five years old (Ewing et al., 2015).

There are various theories on why stereotypes, including face stereotypes, emerge and persist (see Hilton & von Hippel, 1996 for a thorough review). One group of theories suggests that stereotypes emerge from the tension between the complexity of environment around us and our limited cognitive capacity for information processing (e.g., attention, memory). For example, people may be disproportionately aware of salient and distinctive events but under-represent less salient events in memory (Tversky & Kahneman, 1974). Such cognitive limitations may distort the statistical regularities in our mental representation, resulting in cognitive and perceptual biases such as illusory correlations (Hamilton & Gifford, 1976). Another group of theories emphasizes the motivational influences in stereotype formation. A famous example is the self-fulfilling prophecies, which suggest that people may act to make their initially false belief regarding someone or some group come true, regardless of the reality (Snyder et al., 1977).

An important yet under-explored approach is the information sampling perspective (Fiedler, 2000; Fiedler & Juslin, 2006). This view has a unique ecological merit as it preserves the sequential experience sampling nature of almost all real-life social interactions. Moreover, it is practically impossible for people to acquire all relevant information and therefore information sampling has to be selective. In this view, stereotypes arise from making inferences regarding a hypothesis based on selective and potentially biased samples of occurrences in social interactions. It is an established idea that, in social interactions, decisions to make further interactions with someone (and information sampling therein) depend on the valence of the acquired impression of that specific person

[1] Corresponding author.

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society.*

(Lott & Lott, 1972; Montoya & Horton, 2004). Drawing upon this set of ideas, Denrell and colleagues have demonstrated such endogenous sequential sampling processes may underly a number of well-known phenomena in social cognition, including stereotyping and illusory correlations (Denrell, 2005; Denrell & Le Mens, 2011).

We investigate how the endogenous sequential experience sampling process alters the updating of pre-existing face stereotypes in this paper. Previous research suggests that exogenously imposed information can update stereotyped facial trustworthiness to properly guide decision making (Rezlescu et al. 2012). To our knowledge, there is no prior research on the interaction between pre-existing stereotypes and endogenously sampled information. To illustrate, consider the case of playing the Trust Game with a stranger (Berg, et al., 1995, see details below). The investor's first impression of the trustee comes from face stereotypes (Todorov, 2017). Suppose the investor can decide whether or not to invest in the trustee. Such selective experience sampling may produce illusory correlations between facial trustworthiness and actual trustworthiness even when the two attributes are indeed independent (Denrell & Le Mens, 2011).

We test our hypothesis in two experiments using adapted Trust Games (Berg, et al., 1995). In Experiment 1, participants in the role of an investor were asked to make investment decisions based on two attributes of information: an immediately available face image conveying the trustee's facial trustworthiness and the trustee's past performance that had to be *sampled by experience* at the participants' own choice. Participants were asked to sample four of the eight candidate trustees. Consistent with our hypothesis, the participants' trustee choice for experience sampling was highly dependent on the trustee's facial trustworthiness. This asymmetric trustee selection process further increased participants' willingness to invest in trustees with trustworthy faces. In Experiment 2, the two attributes of information were directly *given by description* on one single screen. There was no constraint on information sampling participants were free to look at the information in a self-paced manner. Still, we traced participants' visual information sampling process using an eye-tracker. Both the choice data and eye data suggested that participants regarded past performance as a much more diagnostic attribute and that facial trustworthiness stereotypes had a very limited effect on investment decisions.

In doing so, we showcase a description-experience gap in face stereotyping. This gap is analogous to the well-known description-experience gap in risky choice, where risk-aversion/risk-seeking choice depends on whether the options are directly described in the text or have to be experienced by sampling (Hertwig & Erev, 2009). This gap highlights the essential role of endogenous sequential experience sampling in stereotype persistence.

Experiment 1: Decision from experience

In Experiment 1, we mimicked the naturalistic sequential sampling process of social interactions using a decision-from-experience version of the Trust Game. There were two phases in the task. In the *experience sampling* phase, participants played the Trust Game with some of the trustee players with feedback. Although we constrained the number of times for experience sampling by hand, we believe such a constraint is prevalent in naturalistic settings. In the *decision* phase, participants played the Trust Game with all the eight trustees without feedback. We were particularly interested in how the facial trustworthiness stereotype influenced participants' trustee sampling in the experience sampling phase and willingness to invest in the decision phase.

Methods and Materials

Participants. A total of 60 undergraduate and graduate students (41 females; aged 21.23 ± 2.39) from a public university in Shanghai, China participated in Experiment 1. All participants received a flat payment of 10 Chinese Yuan (1 Chinese Yuan [CNY] was worth 0.15 US dollars at the time of the experiment).

Original Trust Game. Experiment 1 was based on the Trust Game (Berg, et al., 1995). There were two roles in the original game, an investor and a trustee. Participants always played the investor role. At the beginning of the game, participants were endowed with a certain amount of money (50 CNY) and can choose whether or not to invest a part of the money in the trustee (computer generated, not present in the experiment). The invested amount will increase in value by a factor of four. For example, if the investor invests 10 CNY, the amount will be multiplied by a factor of four and the trustee will receive 40 CNY. The trustee then decides whether or not to reciprocate by splitting the bonus with the investor. If the trustee reciprocates (i.e. returning 20 CNY to the investor), the investor ends up earning 10 CNY (i.e., the returned 20 CNY minus the invested 10 CNY) and the trustee earns 20 CNY. If the trustee does not reciprocate, the investor ends up losing 10 CNY while the trustee earns 40 CNY in the game. Therefore, participants' investment decisions depend on their perceived likelihood that the trustee will reciprocate.

Stimuli and design. Experiment 1 used an adapted version of the Trust Game. There were two main screens in the experience sampling phase. On the *profile* screen, we displayed the face images of eight different candidate trustees. These face images were adapted from a previous study (Wang et al., 2019). They included two trustworthy female faces, two trustworthy male faces, two untrustworthy female faces and two untrustworthy male faces (based on trustworthiness ratings from an independent group of 20 participants). On this screen, participants decided which trustee to play the Trust Game with for experience sampling.

We designed two different trustee types in the game, cooperative and defective. Cooperative trustees chose to reciprocate in four of the five rounds, whereas defective trustees chose to reciprocate only in one of the five rounds. Thus, the eight profiles are the products of crossing gender (male versus female), perceived trustworthiness (high versus low) and behavior types (cooperative versus defective).

To further control for potential confounding in other facial features, we created two different versions of profiles so that the trustee type (cooperative versus defective) associated with the two images in each of the four image categories (i.e., trustworthy females, trustworthy males, untrustworthy females and untrustworthy males) flipped. Participants were randomly assigned to one of the two versions.

Procedures. The experiment was programmed on Qualtrics (www.qualtrics.com), and the participants received an anonymous link to complete the online experiment on their personal computers. For data quality control, we requested them to complete the experiment on a computer or a pad (rather than on a mobile phone) within 10 minutes.

After consenting and reading detailed written instructions, the participants were presented with a *profile* screen (Figure 1a) that displayed the eight potential trustee players with their face images (Figure 1). Participants were asked to select four trustees that they wanted to play the game with. Upon the participant's choice, the experiment turned to the *game* screen (Figure 1b), where the participants played the trust with the elected player for five rounds. In each round, participants were endowed with 50 CNY, and were asked to decide how much to invest in the trustee on a slider (ranging from 0 to 50 CNY). Participants got feedback on whether the trustee reciprocated at the end of each round.

After completing the five rounds of the game with a trustee, the program would continue to the new game with the next selected trustee. Participants were asked to play the game with four different trustee players (five rounds each, totaling 20 rounds) for experience sampling.

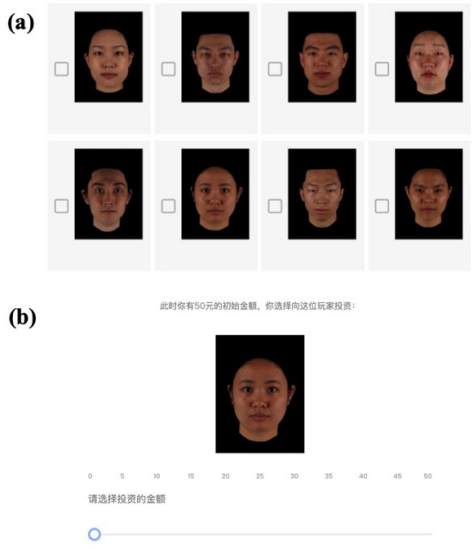

Figure 1: Procedures of the experience sampling phase in Experiment 1. (a) The profile screen presented eight candidate trustees for selection. (b) Game screen, where participants played the Trust Game with a selected trustee for five rounds. Instructions were presented to participants in Chinese (participants' native language).

After the experience sampling phase, participants were asked to play the same trust game with each of the eight trustee players (including four encountered and four unencountered players) for one additional round (totaling eight rounds). No feedback was provided in this decision phase.

Results

Trustee sampling. The first goal of Experiment 1 was to test the effect of facial trustworthiness on the sampling rate. Recall that participants were allowed to sample four out of the eight candidate trustees. Therefore, the average probability of each trustee being sampled was 0.5. Simple summary statistics suggest that trustworthy faces were sampled much more often than untrustworthy faces ($t_{59} = 10.14$, $p < .001$) (Figure 2a). Among those sampled trustees, participants also invested more money in trustees with a trustworthy face than those with an untrustworthy face ($t = 6.31$, $p < .001$ in a multilevel analysis) (Figure 2b). This result confirmed that participants did have an initial bad impression of the seemingly untrustworthy candidates based merely on facial appearance.

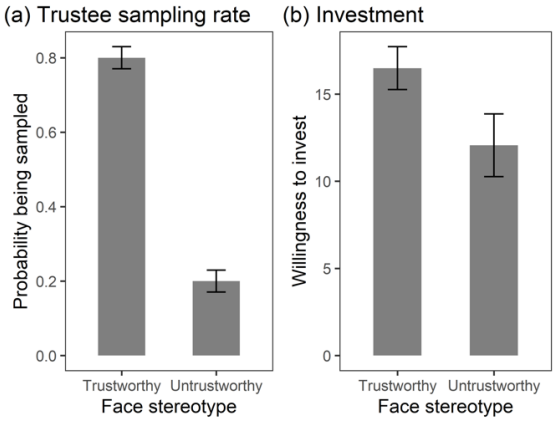

Figure 2: Trustee selection and investment behavior in the experience sampling phase.

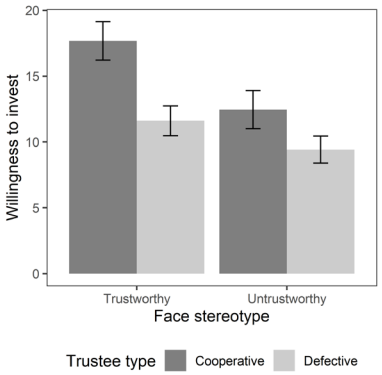

Figure 3: Participants' willingness to invest in all trustees in the decision phase as a function of face stereotypes and trustee types. Error bars represent standard errors.

Investment decisions. The second goal of Experiment 1 was to study how the selective information sampling bias influenced the behavioral consequences of face stereotyping. Participants invested more in trustees with trustworthy faces than in trustees with untrustworthy faces ($F = 20.90$, $p < .001$, $\eta_G^2 = .04$) and invested more in cooperative trustees than in defective trustees ($F = 33.90$, $p < .001$, $\eta_G^2 = .05$). A significant interaction effect between the two independent variables suggests that the main effects were mainly driven by the difference between the cooperative and defective trustees with trustworthy faces ($F = 13.56$, $p < .001$, $\eta_G^2 = .02$, see Figure 3). Because those candidates with untrustworthy faces were rarely sampled, their behavior types (cooperative versus defective) were mostly concealed till the end of the experiment and hence had a much smaller effect on investment decisions.

Experiment 2: Decision from description

Experiment 1 shows that the facial trustworthiness stereotype plays an important role in participants' trustee sampling and hence eventual investment decisions in the decision-from-experience Trust Game. In Experiment 2, we tested the decision-from-description condition in which both the faces and past performance were simultaneously described on one single screen with no sampling constraint. Here, we focus on the primacy of facial trustworthiness on participant's investment decisions as an investor in the Trust Game using both choice data and eye-tracking data.

Methods and Materials

Participants. A total of 71 undergraduate or graduate students (44 female; aged 20.81 ± 2.16) from a public university in Shanghai, China participated in Experiment 2. All participants had normal or corrected to normal vision and none of them had participated in Experiment 1.

Stimuli and design. As in Wang et al. (2022), the face images in Experiment 2 were scraped from an open online employee database of a brokerage company (www.lianjia.com). All face images were anonymized portrait photos displayed in 215px × 300px ellipses. After initial screening, we obtained a total of 400 anonymized face images displayed in 215px × 300px ellipses (200 males and 200 females). We recruited an independent group of participants to rate their facial trustworthiness on a seven-point scale. Each face image received ratings from approximately 20 participants. The average of the ratings was the facial trustworthiness for each face image.

The participants were introduced to the same Trust Game as in Experiment 1. The trustee players' past performance was directly given in the form of donut charts (Figure 4). The green part of the donut ring represented the proportion of reciprocation in the past ten rounds, ranging from 0% to 100%. The past performance for each trustee was randomly generated from a uniform distribution between 0% and 100%.

We created a total of 200 trials from the 400 face images and 400 randomly generated donut charts representing past performance. In each trial, the gender of the two candidates was always the same (either male or female). For one-half of the trials, face images appeared at the top of the screen and the past performance appeared at the bottom. For the other half, past performance appeared at the top and the face images appeared at the bottom.

Procedures. The experiment was run on a Tobii TX300 desktop eye tracker. The eye tracker had a sampling rate of 300Hz and supported a maximum gaze angle of 35°. It was equipped with a built-in 23-inch monitor with a screen refresh rate of 60Hz and a resolution of 1920px×1080px.

Upon arrival, the participants were asked to sit in front of the eye tracker and adjust the chair height to make themselves comfortable looking at the eye tracker screen. They were then asked to complete a Tobii built-in calibration screening with both eyes.

The experiment started after the participants were given detailed instructions about the Trust Game as in Experiment 1. Each trial started with a cross fixation that lasted for 500ms, followed by the display for two trustee profiles, one on the left side and the other on the right side of the screen. Each profile contained two attributes of information: a face image conveying facial trustworthiness and a donut chart indicating past performance. Participants were allowed to look at the information in a self-paced manner. They were asked to imagine they were to play the investor role in the Trust Game and to choose one of the two candidates as the trustee in the game. They indicated their choice with a keypress (F for the candidate on the left side and J for the candidate on the right side).

For each participant, the trials were presented in random order. Participants were allowed to take a self-paced break after every 100 trials.

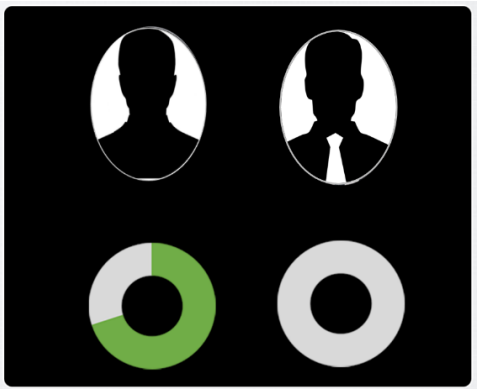

Figure 4: An example trial in Experiment 2. The face images have been replaced by abstract symbols for copyright reasons.

Data pre-processing. To analyze the eye-tracking data, we created four areas of interest (AOI) corresponding to Left Face, Left Performance, Right Face, and Right Performance, respectively. We used circles with a radius of 378px as the AOIs. We chose this radius because it made the AOIs tangential to one another, obtaining the largest number of valid fixations in the eye data. Eye fixations outside of the

AOIs were excluded from the subsequent eye data analysis. Trials that yielded less than two valid fixations were excluded, as no transitions were available. The data from one participant were excluded because of a low gaze sample rate (< 50%, meaning that the eye tracker captured less than half of the gaze data for this participant). After pre-processing, we were left with 13,034 trials from 70 participants.

Results

Attribute primacy in choice data. The primary goal of Experiment 2 was to examine the primacy of facial trustworthiness in investment decisions. Figure 5a shows that at the group level facial trustworthiness had almost no predictive power in accounting for the choice proportions. In stark contrast, Figure 5b shows that the rate of reciprocation in past performance was highly predictive of choice proportions.

To test this more rigorously, we fitted a weighted additive (WADD) model to the choice data to evaluate the extent to which facial trustworthiness, as well as the rate of reciprocation in past performance, determined investment decisions under a hierarchical framework. The main weighted additive choice model, $WADD_{Full}$, took both attributes into consideration:

$$\Pr[Left|\{Left, Right\}]_j = L(\alpha_0 + \alpha_1 \Delta\text{trustworthiness}_j + \alpha_2 \Delta\text{performance}_j) \quad (1)$$

where $L(\cdot)$ is the logistic transformation. $\Delta\text{trustworthiness}_j$ denotes the difference in facial trustworthiness between the left and right profiles in trial j. A positive $\Delta\text{trustworthiness}_j$ indicates that the left profile appeared more trustworthy than the right profile. $\Delta\text{performance}_j$ denotes the difference in past performance between the two profiles. A positive $\Delta\text{performance}_j$ indicates that the left profile had a higher rate of cooperative behavior than the right profile. Both predictors were z-scored in the model such that it is meaningful to compare their coefficients.

To compare the two attributes' relative predictive accuracy, we fit another three sub-models. $WADD_{Performance}$ turned off α_1, $WADD_{Trustworthiness}$ turned off α_2, and $WADD_{Baseline}$ turned off both α_1 and α_2 in the model. As Table 1 shows, $\text{WADD}_{\text{Full}}$, the model that considered both attributes had the best predictive performance, closely followed by $\text{WADD}_{\text{Performance}}$, the model that only used the rate of reciprocation in past performance as the predictor. In contrast, $\text{WADD}_{\text{Trustworthiness}}$, the model that used only facial trustworthiness as the predictor, had the worst fit to the choice data, even below the baseline model when model complexity was taken into account. Again, the model comparison results suggest that facial trustworthiness almost provided no predictive power for participants' binary investment decision when the more diagnostic attribute was directly given in description.

Table 1: Model performance in predicting choice data in Experiment 2.

Model	AIC
$\text{WADD}_{\text{Full}}$	5,289
$\text{WADD}_{\text{Performance}}$	5,344
$\text{WADD}_{\text{Trustworthiness}}$	18,075
$\text{WADD}_{\text{Baseline}}$	18,070

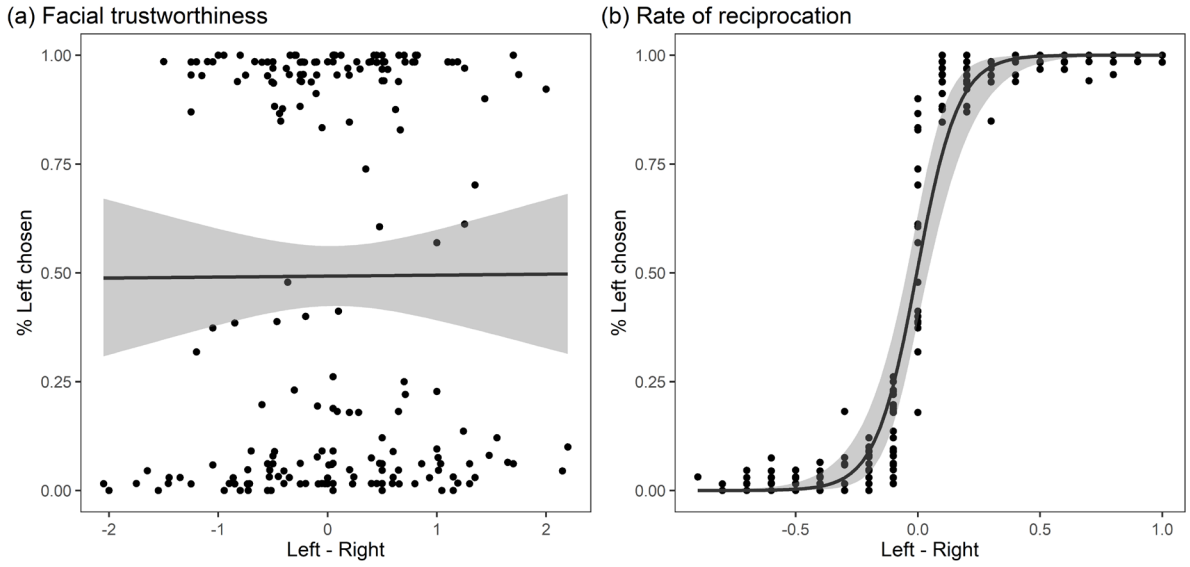

Figure 5: Fitting the choice data to different attributes at the group level. (a) Facial trustworthiness, (b) Rate of reciprocation in past performance. Each data point represents a unique trial.

Attribute primacy in eye-tracking data. Eye movement data provided rich information that reflects how participants prioritize and weigh different attributes of information. We focused on two different types of signals

that reflect attribute primacy. One was that the more important attribute should attract more first fixations. In aggregate, 68.7% of the fixations dwelled in past performance information (i.e., the donut charts). Individually, 78.6% of the participants fixated on past performance more often than on faces during the first fixation in a trial.

The other was that primary attributes attract more attention. It is reasonable to assume that more important attributes would attract more eye fixations in the decision process. In terms of the total number of fixations, in aggregate, participants fixated on past performance 50.6% of the time whereas fixating on faces 49.4% of the time. Although the two numbers appeared almost equal, it is important to note that the faces contained much richer information than the donut charts. In tasks in which the faces are the primary attributes, the proportion of fixations dwelling in faces reached as high as 70% (Wang et al., 2022). Individually, 55.7% of the participants fixated more often on past performance than on faces.

Overall, the eye tracking data also suggest that faces were a less important attribute than past performance when both attributes were simultaneously given with no sampling constraint.

Discussion

This paper studies the persistence of stereotyping in social interactions. In two experiments, we find a description-experience gap in face stereotyping. Using the Trust Game, we study an endogenous information sampling bias that people tend to interact with the seemingly trustworthy trustees based on their facial features in Experiment 1. It is important to note that although we call it a bias, it may be an adaptive heuristic in the face of limited information availability. Experiment 2 shows that when the trustee's actual behavior history is directly given (i.e., the information sampling bottleneck is removed), the face stereotype has a very small effect on participants' willingness to invest in Trust Game. Overall, our study presents a novel description-experience gap and advances our understanding of the cognitive and environmental basis of face stereotyping.

Some recent studies have begun to investigate the information sampling process in face stereotyping. For example, Ma and colleagues have used an information sampling trust game to study informational, social and developmental motives in information sampling for decision making (Ma et al., 2020; 2022). More related to our work, Bai et al. (2021) study the relationship between face stereotyping and information sampling. They find that people do not sample enough information in learning the relationships between facial features and the underlying behavior types and show that such an under-sampling bias undermines the learning outcome in nonstationary environments. In this paper, we argue that even in a stationary environment, the endogenous information sampling bias can distort the perceived relationships between facial features and actual behavior types, and hence amplifies our responses to them.

Perhaps, the most important contribution of this work is that our sampling-based paradigm represents a central aspect of real-life stereotyping problems that have been overlooked in the vast majority of literature on stereotyping. Numerous studies have been conducted to investigate the stereotyping problem from psychological and cognitive perspectives, but little attention has been paid to the interactions between stereotype information sampling. Our results suggest this missing part underlies plenty of the influence of stereotyping on social cognition and behavior. Indeed, our work is motivated by the theoretical demonstration on sampling-based theories of impression formation and beyond, especially those by Denrell and colleagues (Denrell, 2005; Denrell & Le Mens, 2007). Therefore, our work can be extended to other domains with selective experience sampling structures, such as social influence, employment decision making and business ventures.

Our experiments have a few limitations that can be addressed in future research. For example, in Experiment 1, we did not consider learning during the experience sampling phase. We also did not consider other factors that may impact the information sampling processes, including spatial and social distances between different players. Future research can extend the current study design to involve these factors with ease. One lurking issue with our own design was that the stimuli in our two experiments differed in a few manners due to pragmatic constraints. For example, the face images involved in Experiment 2 were different from those in Experiment 1. Further, the two experiments adopted different response modes for dependent variable measurement. While Experiment 1 used a continuous variable (i.e., participants' investment amount) as the dependent variable, Experiment 2 used a binary variable (i.e., participants' decision to invest). Finally, we set the two attributes orthogonal (i.e., zero correlation) in both experiments. The natural environment may be packed with more nuanced regularities and patterns, which could be further explored in future research.

In conclusion, our study provides novel experimental evidence for the sampling-based account for the perseverance of stereotyping and therefore paves a new avenue for testing similar ideas in social cognition and behavior. We did so by describing an essential but overlooked description-experience gap in face stereotyping. By introducing the experience sampling process in the experimental paradigm, we show that the endogenous information sampling bias is responsible for a considerable portion of the behavioral consequences of stereotyping. When this process is eliminated, the perseverance of stereotyping is largely alleviated. We look forward to continually investigating the sampling-based account for social cognition and behavior.

Acknowledgements

The authors thank four anonymous reviewers for helpful comments. This research was supported by the National Natural Science Foundation of China [grant number:

72101156, 71901095] and Shanghai Pujiang Program [grant number: 2020PJC102].

References

Bai, X., Uddenberg, S., Labbree, B. P., & Todorov, A. (2021). Insta-learn: Face Stereotypes Emerge and Persist through Insufficient Statistical Learning. https://doi.org/10.31234/osf.io/te3s4

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. *Emotion, 6*(2), 269-278.

Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and Economic Behavior*, *10*(1), 122-142.

Denrell, J. (2005). Why most people disapprove of me: experience sampling in impression formation. *Psychological Review*, *112*(4), 951-978.

Denrell, J., & Le Mens, G. (2011). Seeking positive experiences can produce illusory correlations. *Cognition*, *119*(3), 313-324.

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology, 24*(3), 285-290.

Ewing, L., Caulfield, F., Read, A., & Rhodes, G. (2015). Perceived trustworthiness of faces drives trust behaviour in children. *Developmental Science, 18*(2), 327-334.

Fiedler, K. (2000). Beware of samples! A cognitive-ecological sampling approach to judgment biases. *Psychological Review*, *107*(4), 659-676.

Fiedler, K., & Juslin, P. (Eds.). (2006). *Information Sampling and Adaptive Cognition*. Cambridge University Press.

Hamilton, D. L., & Gifford, R. K. (1976). Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgments. *Journal of Experimental Social Psychology*, *12*(4), 392-407.

Hertwig, R., & Erev, I. (2009). The description–experience gap in risky choice. *Trends in Cognitive Sciences*, *13*(12), 517-523.

Hilton, J. L., & Von Hippel, W. (1996). Stereotypes. *Annual Review of Psychology*, *47*(1), 237-271.

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin, 126*(3), 390.

Lott, A. J., & Lott, B. E. (1972). The power of liking: Consequences of interpersonal attitudes derived from a liberalized view of secondary reinforcement. In *Advances in Experimental Social Psychology*. Academic Press.

Ma, I., Sanfey, A. G., & Ma, W. J. (2020). The social cost of gathering information for trust decisions. *Scientific Reports*, *10*(1), 1-9.

Ma, I., Westhoff, B., & van Duijvenvoorde, A. C. K. (2022). Uncertainty about others' trustworthiness increases during adolescence and guides social information sampling. *Scientific Reports*, *12*(1), 7634.

Maestripieri, D., Henry, A., & Nickels, N. (2017). Explaining financial and prosocial biases in favor of attractive people: Interdisciplinary perspectives from economics, social psychology, and evolutionary psychology. *Behavioral and Brain Sciences, 40*, e19.

Montoya, R. M., & Horton, R. S. (2004). On the importance of cognitive evaluation as a determinant of interpersonal attraction. *Journal of Personality and Social Psychology, 86,* 696–712.

Snyder, M., Tanke, E. D., & Berscheid, E. (1977). Social perception and interpersonal behavior: On the self-fulfilling nature of social stereotypes. *Journal of Personality and Social Psychology*, *35*(9), 656-666.

Rezlescu, C., Duchaine, B., Olivola, C. Y., & Chater, N. (2012). Unfakeable facial configurations affect strategic choices in trust games with or without information about past behavior. *PloS one*, *7*(3), e34293.

Todorov, A. (2017). Face value. In *Face Value*. Princeton University Press.

Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science*, *185*(4157), 1124-1131.

Van't Wout, M., & Sanfey, A. G. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition, 108*(3), 796- 803.

Wang, H., Han, C., Hahn, A. C., Fasolt, V., Morrison, D. K., Holzleitner, I. J., ... & Jones, B. C. (2019). A data-driven study of Chinese participants' social judgments of Chinese faces. *PloS one*, *14*(1), e0210315.

Wang, H., Ma, J., & He, L. (2022). Value-directed information search in partner choice. *Judgment and Decision Making, 17*(6), 1287-1312.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100- ms exposure to a face. *Psychological Science, 17*(7), 592-598.